



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2013

Standardization of baseline and additionality determination under the CDM

Hayashi, Daisuke ; Michaelowa, Axel

Abstract: Clean Development Mechanism (CDM) project developers have long complained about the complexities of project-specific baseline setting and the vagaries of additionality determination. In response to this, the CDM Executive Board took bold steps towards the standardization of CDM methodologies, culminating in the approval of guidelines for the establishment of performance standards in November 2011. The guidelines specify a performance standard stringency level for both baseline and additionality of 80% for several priority sectors and 90% for all other sectors. However, an analysis of 14 large-scale CDM methodologies that use performance standard approaches challenges this top-down approach to the performance standard design. An appropriate performance standard stringency level strongly depends on sector and technology characteristics. A single stringency level for baseline and additionality determination is appropriate only for greenfield projects, but not for retrofit ones. Overly simple, highly aggregated performance standards are unlikely to ensure high environmental integrity, and difficult questions regarding stringency and updating frequency will eventually have to be addressed on a rather disaggregated level. A careful balance between data requirements and the practicability of performance standards is essential because the heavy data requirements of the existing performance standard methodologies have been the key barrier to their actual implementation. Policy relevance CDM regulators have been pushed by many stakeholders to standardize baseline setting and eliminate project-specific additionality determination. At first glance, performance standards seem to provide the perfect solution for both tasks. However, a one-size-fits-all political decision – e.g. the average of the top 20% performers as enshrined in the Marrakech Accords – is inappropriate. Substantial disaggregation of performance standards is required both technologically and geographically in order to limit over- and under-crediting and close loopholes for non-additional projects. As a lack of reliable and complete data has been and will be a key bottleneck for the development of performance standards, international support for data collection will be indispensable, but costly, and time-consuming. Empirically driven, techno-economic assessments of performance standard stringency levels must be the central task of the future work on standardized methodologies, and should not be sidelined by perceived needs of policy makers to take bold decisions under time pressures.

DOI: <https://doi.org/10.1080/14693062.2013.745114>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-77129>

Journal Article

Accepted Version

Originally published at:

Hayashi, Daisuke; Michaelowa, Axel (2013). Standardization of baseline and additionality determination under the CDM. *Climate Policy*, 13(2):191-209.

DOI: <https://doi.org/10.1080/14693062.2013.745114>

**Standardization of baseline and additionality determination
under the Clean Development Mechanism**

Daisuke Hayashi^{1,*} and Axel Michaelowa^{1,2}

¹ Department of Political Science, University of Zurich, Affolternstrasse 56, 8050 Zurich, Switzerland

² Perspectives, Klosbachstrasse 2, 8032 Zurich, Switzerland

* Corresponding author: daisuke.hayashi@pw.uzh.ch

Abstract: Clean Development Mechanism (CDM) project developers have complained for many years about the complexities of project-specific baseline setting and the vagaries of additionality determination. In response to this, the CDM Executive Board took bold steps towards standardization of CDM methodologies, culminating in approval of guidelines for the establishment of performance standards in November 2011. The guidelines specify performance standard stringency for both baseline and additionality at a level of 80% for several priority sectors and 90% for all other sectors. However, an analysis of 14 large-scale CDM methodologies using performance standard approaches challenges this top-down approach to the performance standard design. An appropriate level of performance standard stringency strongly depends on sector and technology characteristics. A single stringency level for baseline and additionality determination is appropriate only for greenfield projects, but not for retrofit ones. Overly simple, highly aggregated performance standards are unlikely to ensure high environmental integrity, and difficult questions regarding stringency and updating frequency will eventually have to be addressed on a rather disaggregated level. A careful balance between data requirements and practicability of performance standards is essential because heavy data requirements of the existing performance-standard methodologies have been the key barrier to the actual implementation.

Policy relevance: CDM regulators have been pushed by many stakeholders to standardize baseline setting and eliminate project-specific additionality determination. At first glance, performance standards seem to provide the perfect solution for both tasks. However, a one-size-fits-all political decision like the average of the top 20% performers enshrined in the Marrakech Accords is inappropriate. Substantial disaggregation of performance standards is required both technologically and geographically in order to limit over- and under-crediting and close loopholes for non-additional projects. As lack of reliable and complete data has been and will be a key bottleneck for the development of performance standards, international support for data collection is indispensable; it will be costly and time-consuming. Empirically driven, techno-economic assessment of performance standard stringency levels must be the central task of the future work on standardized methodologies, and should not be sidelined by perceived needs of policymakers to take bold decisions under time pressure.

Keywords: Performance standard, baseline, additionality, Clean Development Mechanism (CDM), Kyoto Protocol

1. Introduction

The objectives of the Clean Development Mechanism (CDM), as set out in the Kyoto Protocol, are twofold: (i) to assist developing countries (non-Annex I countries) to achieve sustainable development whilst at the same time (ii) to assist industrialized countries (Annex I countries) in achieving compliance with their quantified emissions targets in a cost-effective manner (UNFCCC, 1998). While the achievement of the latter target is without doubt given the huge numbers of CDM projects registered to date and the volume of issued CDM credits (Certified Emission Reductions, CERs) reaching over 1 billion by the end of 2012, the regional distribution of CDM projects is very uneven and some sectors such as transport and demand-side energy efficiency improvement have essentially been sidelined (URC, 2012). Due to these shortcomings an increasing number of stakeholders have called for a reform of the CDM. The cumbersome procedure of baseline setting and additionality demonstration has been one of the major problems identified in these reform calls (e.g., Hayashi and Michaelowa, 2007; IETA, 2009; UNFCCC, 2010).

CDM baseline and monitoring methodologies are usually developed by project developers and then checked by the Methodologies Panel of the CDM Executive Board (CDM EB); the rejection rate is high (from a minimum of 39% in 2004, it increased to a peak of 58% and in June 2012 stood at 56%). Due to this bottom-up nature of methodology development, project developers have naturally preferred designing methodologies around project-specific factors to proposing standardized approaches that can be used by their competitors. Although there have been attempts by the CDM regulators to consolidate approved methodologies into a more user-friendly, broadly applicable methodology (by June 2012, 19 single methodologies had been combined into 11 consolidated methodologies), the great complexity in the current CDM methodologies is often criticized as one of the key deficiencies of the CDM.

Determination of project additionality, i.e., whether a CDM project is mobilized through the CDM revenue and thus not “business-as-usual” or the continuation of current practice (see Gillenwater (2011) for a compilation of additionality definitions), has been a heavily contested subject. In 2005, the CDM EB introduced an “additionality tool” which assesses whether the project is financially less attractive than a realistic alternative, or whether it faces prohibitive barriers. For a comprehensive history of additionality determination rules see Michaelowa (2009).

Against this background, greater standardization is currently seen as a key means of improving the efficiency of the CDM. CDM methodologies would become more streamlined and objective, and participation of currently underrepresented host countries would be encouraged by removing the methodological barriers and shifting the cost of methodology development and data collection from individual project developers to a centralized coordinating body (Hayashi et al., 2010). At the international climate negotiations standardization of baselines has recently become a relevant topic. The sixth session of the Conferences of Parties serving as the Meeting of the Parties (COP/MOP6) in December 2010 in Cancun, Mexico requested the CDM EB to initiate its work on development of standardized baselines (UNFCCC, 2011a), which was then vigorously pursued by the CDM EB during 2011. In September 2012, the CDM EB established a procedure for submission and consideration of standardized methodologies. This is essentially a bottom-up process in which countries, project participants, international industry organizations or admitted observer organizations may propose a standardized methodology through the Designed National Authority (DNA) of a country for which the standardized methodology is proposed (UNFCCC, 2012).

Standardization can be achieved through two main routes (UNFCCC, 2010). The first route would be performance standards such as emissions or energy intensity denominated in emissions or energy per unit of output and determined by a peer group of similar plants or installations. The second, less bold route

would be the use of default values, i.e., emission factors and emission reduction values used as inputs to the calculation of baseline, project or leakage emissions. A combination of these two approaches is also possible. For instance, some default values can be used in establishing a performance standard in order to reduce monitoring costs of the performance standard approach.

While the default factor approach has already been applied in a number of CDM methodologies, the performance standard approach is a methodological innovation recently achieved under the CDM. The 65th meeting of the CDM EB in November 2011 defined a performance standard specifying automatic additionality of projects as well as baseline emission factors (UNFCCC, 2011b). The guideline applies two types of performance standards: (i) technology or practice standard, and (ii) emission standard. However, they cannot be applied “off the shelf” and need to be submitted as a standardized methodology by host country DNAs, underpinned by empirical data (UNFCCC, 2012).

Technology or practice standards are applied for methane destruction or avoidance projects. These projects are deemed additional if they destruct or avoid methane emissions more than what is legally required. A baseline of methane destruction projects is defined as the regulated level of methane destruction, while the most commonly used disposal and treatment method sets a baseline for methane avoidance project.

Emission standards are applied to projects for fuel or feedstock switch, and technology switch with or without change of energy sources (including energy efficiency improvement). This approach sets a baseline and demonstrates additionality based on a pre-defined level of emission performance (in tCO₂e/unit of output). Energy for households, energy generation in isolated systems and agriculture are defined as “priority sectors” in the determination of emission performance thresholds. The performance standards are set at 80% for the priority sectors, and at 90% for all other sectors, of an emission performance curve

of relevant technologies. That is, a baseline emission factor is determined by the emission factor of the fuel, feedstock or technology sitting at the 80% (90%) level of an emission performance curve. Figure 1 illustrates an emission performance curve of electricity generation technologies. In this example, a baseline would be electricity generation using natural gas for the 80% level, and hydropower generation for the 90% level.

[Insert Figure 1 here]

Figure 1: Emission performance standard for electricity generation as per CDM EB guideline

Source: Authors.

Additionality of these project types is demonstrated if the following two conditions are met: (i) the implementation of a mitigation measure results in an emission factor that goes beyond an additionality performance standard set at 80% (90%) of the emission performance curve, and (ii) the mitigation measure faces barriers for implementation or it is less commercially attractive than all other technologies used to produce outputs with an emission performance lower than the additionality performance standard. The measures fulfilling these conditions are added to a positive list, where additionality of the listed measures is automatically granted.

This article focuses on the emission and energy performance standard approaches, and assesses the potential of the greater use of such standards as a means to standardize procedures for CDM baseline setting and additionality demonstration. Section 2 explains principles of CDM methodologies in order to set judgment criteria for performance standard approaches. Section 3 reviews existing literature to summarize performance standard design. Based on this conceptual framework, section 4 conducts a systematic review of performance standard approaches used in existing CDM methodologies. The actual

application of these methodologies to CDM projects is analyzed in section 5. Finally, section 6 concludes with lessons for further standardization of CDM methodologies using performance standard approaches.

2. Principles of CDM methodologies

Before starting a discussion on standardized approaches to baseline setting and additionality demonstration of CDM projects, it is necessary to reflect on the following question: “What makes a good CDM methodology?” The United Nations Framework Convention on Climate Change (UNFCCC) has not published any specific guidance on the principles of standardized approaches, but there exist two official documents that explain fundamental requirements of CDM methodologies.

The first is a decision taken by the 1st session of COP/MOP in 2005 (UNFCCC, 2006). The COP/MOP decision stipulated that CDM methodologies should be broadly applicable, promote consistency, transparency and predictability, and provide rigor to ensure that emission reductions are real and additional (UNFCCC, 2006). The second is the CDM EB’s 49th meeting report published in 2009, which specified the following criteria for improving approved methodologies that had not been widely utilized by CDM project developers: objectivity, applicability, usability and consistency (UNFCCC, 2009). Table 1 summarizes the above principles of CDM methodologies, with our own account of how these principles should be interpreted in the context of standardized approaches.

[Insert Table 1 here]

These principles are generally applicable for both project-specific and standardized CDM methodologies. However, the principle of environmental integrity for standardized methodologies requires a different

interpretation than for conventional methodologies. While the latter are designed to ensure that emission reductions of individual CDM projects are real and additional, the former should meet this requirement only at an aggregate level, but not at an individual project level. A performance standard is by nature designed to serve multiple projects, so it might as well over- or under-credit emission reductions achieved by each of these projects. But the overall environmental integrity can be maintained as long as the total amount of over-crediting and under-crediting is kept in balance. It should be kept in mind that under-crediting can take place only if there are genuinely additional projects that generate fewer credits than they would have generated under a project-specific baseline. But there might be cases where all projects generating credits under a performance threshold are non-additional. Let us assume that, in a specific host country, supercritical coal power plants beat the threshold despite being commercially attractive. Thus, all these plants are registered as CDM projects, and given limited demand for credits, no other CDM projects for power generation are submitted. In this situation, the over-crediting of the supercritical coal power projects is not balanced by any under-crediting.

It is also important to note that a key strength of performance standards lies in their ability to accommodate a series of mitigation measures implemented together. In such a case, it is often challenging to establish clear causality between the overall emission reductions and every single measure implemented because these measures often interact with each other. Accordingly, a clear distinction between additional and non-additional measures is often not possible in practice. Therefore, performance standards need to assess a baseline and additionality in a holistic manner by conservatively aggregating the causality at a system level.

3. Concept of performance standards

Performance standards are generally defined as a “comparison of performance against peers based on a set of criteria”. To date, many industries have gained experience with performance standards, mainly as a management tool for identifying potential for improvement in operation (Neelis et al., 2009). On the other hand, there is yet limited experience with the application of performance standards for international comparison of GHG emission performance (for a review of existing initiatives, see Hayashi et al., 2010). The analysis in this section focuses on the following key design issues of performance standards in the area of climate change mitigation identified in the literature: (i) performance indicators, (ii) level of aggregation, (iii) stringency, (iv) data requirements, and (v) updating procedures (Broekhoff, 2007; Lazarus et al., 1999 and 2000; Hayashi et al, 2010; UNFCCC, 2010).

3.1 Performance indicators

There are different ways to define indicators for comparison of GHG emission performance against peers. The following three types of indicators are proposed or implemented most widely in climate change mitigation:

- **Emission or energy standard:** An emission or energy standard is expressed with (i) a numerator for evaluating climate impacts (e.g., GHG emissions or energy consumption), and (ii) a denominator that refers to a functional category of the output provided (e.g., electricity generation) (IETA, 2009; Hayashi et al., 2010; Lazarus et al., 1999 and 2000). One can then define a threshold value of specific emissions or specific energy consumption for baseline setting and/or additionality demonstration of a CDM project. Such a threshold value can be determined by performance of peers and/or GHG emission or energy efficiency standards implemented in a relevant jurisdiction.

- **Technology or practice standard:** Normative specifications on technology or practice are used to determine baseline emissions or assess additionality of a CDM project (IETA, 2009; Hayashi et al., 2010; Lazarus et al., 1999 and 2000). For example, such specifications may be based on the best available technology or practice in a given geographical boundary (e.g., use of an anaerobic digester for the control of methane emissions from dairy livestock). The prescribed technology or practice can be used for baseline setting and/or additionality demonstration.
- **Market penetration rate:** A certain threshold on market penetration of a mitigation measure is used for additionality demonstration of a CDM project (Kantha et al., 2005, Hayashi et al., 2010). The rationale for this approach is that emerging technologies (i.e., penetration rate is lower than a threshold) are likely to require some type of support in order to compete with other technologies available in the market. Hence, they can be deemed additional and qualify for receiving financial support through the CDM. For instance, CDM projects for distribution of compact fluorescent lamps may be deemed additional until a market penetration rate of the technology exceeds 5%.

3.2 Level of aggregation

The grouping of various types of potential projects into a single category with a corresponding single baseline is the defining aspect of performance standards. A single performance standard that is based on a sector-wide aggregation would provide a strong incentive for investing in the least carbon-intensive technologies available in the sector (Lazarus et al., 1999). Once established, such a highly aggregated performance standard could significantly reduce transaction costs of baseline setting and additionality demonstration.

However, depending on sector characteristics, a certain degree of disaggregation is necessary for designing performance standards. Four key dimensions of aggregation are: (i) process, (ii) product, (iii) time, and (iv) space. First, the process dimension asks whether performance standards are differentiated by technology or process. Second, the product dimension analyses whether the product or service for performance comparison should be further disaggregated. Third, the temporal dimension assesses the vintage of peers for performance comparison. Lastly, the spatial dimension determines the geographical boundary in which the peers are located (Hayashi et al., 2010).

The following example gives an insight into the process aggregation. Suppose a power sector in which 20% of its total power generation is based on renewables and 80% based on coal (i.e., roughly about China's electricity mix in 2005, see IEA, 2007a). If a sector-wide performance standard were set and a stringency level most commonly used in CDM methodologies is applied (i.e., the average of the top 20% performers), the performance standard would be zero emissions and virtually no project would be eligible for the CDM. In order to provide incentives for emission reductions for the dominant coal power plants, performance standards need to be differentiated by feedstock (renewables vs. coal).

As for the product aggregation, consider a pulp and paper sector that produces a wide range of products such as wood pulp, boxboard, newsprint, finished paper. These products have very different energy requirements (IEA, 2007b). If they were perfectly substitutable commodities, a sector-wide performance standard could be set across the product range. However, there are distinguished market demands for these products because each of them has different product properties. Therefore, product-specific performance standards would be more appropriate than a single performance standard set across the product range.

The temporal aggregation plays a role when a sector consists of facilities (e.g., plants, buildings) with varying vintage. Newer facilities tend to employ more modern, climate-friendly technologies or practices

than older ones do. In addition, environmental regulations are commonly set stricter for newer facilities than for older ones (e.g., some building efficiency standards are applicable only to new construction of buildings). These effects combined, newer facilities tend to perform better in terms of GHG emissions or energy consumption. Thus, the choice of peers for performance comparison needs to account for this vintage effect. This can be made by differentiating performance standards by new and existing facilities, and if necessary, by further disaggregating the existing facilities by vintage class (Hayashi et al., 2010).

Performance standards need to be established for a relevant spatial boundary because local parameters can have significant influence on GHG emission performance as well as on the costs and potential for emission reductions. Particularly influential local parameters include the availability, costs, and carbon intensity of energy inputs and feedstock (Lazarus et al., 1999; Hayashi et al., 2010). In addition, scope of trading of interim products may become a decisive factor (Hayashi et al., 2010). In many industrial sectors, production plants use interim products purchased from third parties for producing final products (e.g., in the cement sector, it is common to import clinker, an interim product, for producing cement, a final product). The emission or energy performance of the interim product influences the one of the final product. Therefore, if there is a large difference in the emission or energy performance of the interim products traded in different regions, it is necessary to account for the regional difference in establishing performance standards.

3.3 Stringency

There are two issues related to stringency of performance standards. First, it needs to be determined which level on a performance curve should be used as a performance threshold. Second, a decision needs to be taken whether the same level of stringency should be used for baseline setting and additionality demonstration or different thresholds be used.

3.3.1 Stringency level and performance distribution

A performance standard needs to be set at a stringency level that strikes an overall balance between over-crediting and under-crediting of mitigation efforts in a sector, i.e., at the level that ensures the environmental integrity. A performance standard that is set more stringent than this level risks the unnecessary exclusion of potential emission reductions that are real and additional. This would mean that an increase in stringency beyond that level does not improve environmental integrity of the CDM, but only reduces the amount of emission credits generated. An overly stringent performance standard also fails to provide incentives for improvement to the worst-performing emitters that lag far behind the stringent performance standard (Hayashi et al., 2010). On the other hand, a too lenient performance standard will give too many credits for phantom reductions (Broekhoff, 2007).

An appropriate level of stringency likely differs by sector, mainly depending on a distribution curve of emission or energy performance of entities in the sector (Hayashi et al., 2010). For instance, if a sector is very homogeneous and has a narrow, normal distribution of emission or energy performance, a performance standard can reasonably be set close to the average carbon intensity level (case (a) in Figure 2). This is because the average performance level has the best chance of balancing the over- and under-crediting. On the other hand, the average level would not represent the most likely counterfactual scenario if a sector is very heterogeneous and has a largely skewed or multimodal distribution of emission or energy performance (case (b) in Figure 2). In such cases, it may be advisable to set different levels of stringency for different segments of the performance distribution.

[Insert Figure 2 here]

Figure 2: Stringency level for normal and multimodal performance distribution

Source: Authors.

It is easier to find a stringency level that balances over- and under-crediting if sectors are highly disaggregated (e.g., clinker production as a sub-sector of the cement sector compared to the entire cement sector). The higher the aggregation level of a sector, the more heterogeneous its emission or energy performance tends to be due to, e.g., increasing diversity in technology or practice used in the sector. Thus, stringency and aggregation aspects closely interact with each other in determining the environmental integrity of a performance standard.

3.3.2 Single or dual stringency level for baseline and additionality

Another consideration is whether the same or different stringency levels should be applied to baseline setting and additionality demonstration (Broekhoff, 2007). Either a single stringency level can be used for baseline setting and additionality demonstration (case (a) in Figure 3), or two different levels can be used (case (b) in Figure 3). In case of the latter, the threshold for additionality demonstration is set more stringently than for baseline setting in order to reduce the risk of crediting non-additional emission reductions.

[Insert Figure 3 here]

Figure 3: Single or dual stringency level for baseline and additionality

Source: Authors.

One of the key considerations in choosing a single or dual stringency level is whether a project involves construction of new facilities (greenfield project) or retrofit of existing facilities (brownfield project) (Hayashi et al., 2010). To understand this point, it is important to keep in mind that a baseline can

determine what is business-as-usual or not only if the baseline represents a counterfactual scenario that is most economically attractive after excluding alternative scenarios that face significant barriers for implementation (adapted from Michaelowa, 2005). If a performance standard is set to satisfy this condition, a single stringency level can be used for baseline setting and additionality demonstration.

The condition is far more likely to be met by a greenfield project than a brownfield one. At the time of new facility construction, investment decisions are widely open. Thus, the project developer would likely choose an option that does not face significant implementation barriers and provides the highest economic return on his investment. This logic is reflected in the baseline approach 48.b of the Marrakech Accords, which sets a baseline as “emissions from a technology that represents an economically attractive course of action, taking into account barriers to investment” (UNFCCC, 2001). This approach has been used as a de-facto standard for baseline setting for greenfield projects under the CDM.

In case of brownfield projects, a baseline is commonly set at a historical level of GHG emissions. The historical emissions level does not necessarily correspond to the most economically attractive course of action because, e.g., the existing technology or practice employed by the project may be an obsolete one. If there are business-as-usual mitigation measures available to the existing facility, the use of its historical emissions level for additionality demonstration would produce phantom emission reductions. For this reason, it makes sense to apply a separate, more stringent performance standard for additionality demonstration of brownfield projects.

The above discussions show the fundamental importance of sector and project characteristics in determining stringency of baseline and/or additionality performance standards. A one-size-fits-all approach is thus unlikely to be a good practice for setting stringency.

3.4 Data requirements

Data need to fulfill certain criteria regarding accuracy, confidentiality, relevance and completeness (adapted from CSI, 2011). They need to be sufficiently accurate to provide confidence in performance standard calculation and need to be handled in a way that does not infringe confidentiality of data suppliers, e.g., by displaying data analysis results in an aggregate manner. The relevance of the data is high if they contain all parameters that are required to calculate performance standards. The data also need to be complete, or cover a sufficient share of the sector output, so that a cohort of peers for performance comparison is representative of the entire sector (adapted from Hayashi et al., 2010).

In principle, performance standards can be established using either empirical data or model projections that may include simulations of technology performance over time (Lazarus et al. 1999 and 2000; Murtishaw et al., 2006; Hayashi et al., 2010). Lazarus et al. (2000) and Murtishaw et al. (2006) argue against the use of projection-based data because projections may be more susceptible to deliberate gaming, and they tend to depend on a number of subjective, and often hidden, assumptions. Another key aspect of data requirements is a data timeframe. Based on an empirical analysis of the power sector data from California, the US and Guatemala, Murtishaw et al. (2006) show that using data over five years is a good rule of thumb for setting performance standards in the power sector, and a shorter data period would make performance standards a poor predictor of the average emission rates of subsequent activity.

A performance standard normally requires more extensive data than a conventional, single-project methodology does. Therefore, data availability is one of the most critical issues for practicability of performance standards. The data requirements, in terms of both quantity and quality, need to be balanced against data availability. Ideally, data would be based on actual operating data. However, if such data is not obtainable, an alternative, less accurate option is to substitute “nameplate” parameters based on

manufacturer's specifications for actual operating data. The third tier of data source is the use of default factors (Lazarus et al., 2000). The data should ideally be collected in a concerted manner because data requirements of performance standards are often too heavy for a single project developer. Such efforts could be coordinated by host country institutions (e.g., government, industry association) and/or international institutions with relevant expertise (e.g., UN agency, multilateral development bank). In some cases, collaboration with industry associations such as World Steel or the Cement Sustainability Initiative that have already collected large amounts of data will reduce data collection efforts; of course data quality needs to be assessed by independent auditors.

3.5 Updating procedures

Performance standards need to be updated periodically to reflect changing economic, social, technological, and environmental conditions. Performance standards can be updated by regularly collecting data from a cohort of peers, or based on a pre-defined factor for autonomous improvement in performance indicator (Hayashi et al., 2010).

In defining updating procedures, it is important to keep in mind that frequent updating results in higher monitoring costs and lower investor certainty (Lazarus et al., 2000). It is hence essential to investigate how frequently a performance indicator evolves over time. A sector with frequent changes in emission or energy performance (e.g., a rapidly growing sector with frequent technological improvements) would require more frequent updating than a sector with stable performance over time.

In addition, it is not necessary to regularly update all parameters used for the performance standard setting. For instance, parameters with only minor influence on the entities' performance can be updated less frequently than ones with major influence. Also, some parameters change more frequently than others.

Therefore, updating frequency of each parameter could be differentiated in order to reduce the updating cost.

4. Performance standards applied in large-scale CDM methodologies

This section analyzes large-scale CDM methodologies according to the key design issues of performance standards discussed in section 2.¹ As of March 2012, there were 13 approved large-scale CDM methodologies that apply performance standards. In addition, one rejected large-scale methodology (NM0302) is included in the analysis because it applies a novel performance standard approach (see Table 2).

[Insert Table 2 here]

4.1 Performance indicators

The most common type of performance standards is the emission or energy standard approach, using specific GHG emissions or energy consumption as a performance indicator. The performance standard is mostly used for baseline determination (AM0030, AM0037, AM0059, AM0063, AM0070, AM0091, ACM0002, ACM0013, NM0302). But its application to additionality demonstration has emerged in recent methodologies (AM0070, AM0091, NM0302).

¹ Recently, several small-scale methodologies have been introduced that also apply performance standards, but treatment of these methodologies would go beyond the scope of this paper.

The technology or practice standard approach is applied in a few methodologies for baseline setting (AM0067, ACM0005, ACM0015, ACM0018). The standard establishes a baseline threshold level of technological performance or practice (e.g., “power generation efficiency” in ACM0018).

The market penetration approach is used exclusively for additionality demonstration (AM0014, ACM0005). It defines a market penetration rate below which a project is automatically deemed additional. The market penetration rate is measured either by market share (e.g., “the fraction of total sales of blended cement in the host country” in ACM0005), by installed capacity (e.g., “installed cogeneration capacity” in AM0014), or by the number of installed facilities (e.g., “installed number of cogeneration plants” in AM0014).

4.2 Level of aggregation

4.2.1 Process aggregation

Process aggregation is observed in about half of the reviewed methodologies (AM0030, AM0059, AM0067, AM0091, ACM0002, ACM0013). In earlier methodologies, process disaggregation was not very common, unless there was already an established database to enable such disaggregation (e.g., disaggregation by smelter technology in the aluminum sector, see AM0030 and AM0059).

However, there is an increasing tendency in recent methodologies towards detailed process disaggregation. For instance, AM0067 sets a performance standard per type of transformer, which is defined by capacity and transformation ratio of the transformer. Furthermore, ACM0013 disaggregates the power sector by the type of fuel used for power generation, load category and plant capacity.

4.2.2 Product aggregation

As most of the CDM methodologies are designed for a specific technology that produces a distinctive, homogeneous output, detailed product disaggregation is not very common so far. For instance, AM0030 and AM0059 are applicable to the primary aluminum production process. The performance standard applied in these methodologies use the amount of primary aluminum production, which excludes secondary aluminum.

Detailed product aggregation is applied when a methodology attempts to accommodate a wider range of products. AM0070 is applicable to manufacturing of energy-efficient refrigerators. As different refrigerator models have different energy performance, its performance standards are differentiated by storage volume class and design (“direct cool” or “frost free”) of refrigerators.

AM0091 is applicable to implementation of energy-efficiency measures and/or fuel switching in the construction of new building units (residential, commercial and/or institutional building units). As the methodology has very wide applicability, detailed product disaggregation is necessary for ensuring reasonable accuracy in performance standard setting. The methodology uses gross floor area of a building unit (m²) as a denominator of its performance indicator. Its performance standards are differentiated by building type, size, height and operating hours as well as climate conditions in which buildings are located.

4.2.3 Temporal aggregation

The majority of the methodologies set temporal thresholds for the choice of peers for comparison. The threshold is typically set as the “most recent five years” (AM0037, AM0067, AM0091, ACM0013, ACM0018, NM0302). This is likely referenced to the requirement of the baseline approach 48.c of the

Marrakech Accords, which determines a baseline as “the average emissions of similar project activities undertaken in *the previous five years*, in similar social, economic, environmental and technologies circumstances, and whose performance is among the top 20 per cent of their category” (UNFCCC, 2001; *italic added by the authors*).

However, there are also other variants such as “the most recent 10 years” (CO₂ recovery in AM0063), “five most recently built plants, or a set of recently built plants that comprise 20% of the total power production in the grid” (for the calculation of build margin in ACM0002), and “the most recent year” (efficient refrigerators in AM0070). Performance standards applied to facilities or products with long lifetimes of capital stock tend to set longer timeframes for the temporal threshold.

4.2.4 Spatial aggregation

The most commonly used spatial boundary is the host country (AM0014, AM0037, AM0063, AM0070, ACM0005), or the power grid system if a project concerns power generation (AM0067, ACM0002, ACM0013). Other spatial boundaries are used considering the influence of local parameters on emission or energy performance. AM0091 uses a municipality as a spatial boundary because local parameters (e.g., climate conditions) can have significant influence on building emission or energy performance. NM0302 proposes the use of a global boundary for determining an energy-efficiency performance standard for new cement plants because newly constructed cement plants have comparable efficiency levels around the world. On the other hand, it applies a regional boundary for parameters sensitive to local conditions (e.g., fuel mix, cement blending ratio).

Several methodologies define a spatial boundary based on the geographical area in which a product is traded. AM0030 and AM0059 use a global boundary for the primary aluminum production, ACM0015

sets its boundary as a 200-km radius from the cement plant, and AM0037 uses the host country if a chemical product is traded regionally, or all countries if it is globally traded.

Some of the above methodologies define a minimum sample size for the calculation of performance standards, and require the boundary to be expanded until the sample size exceeds the minimum requirement (AM0037, AM0063, AM0091, ACM0013).

4.3 Stringency

The dominant choice of stringency level is “the average of the top 20% performers” (AM0037, AM0059, AM0067, AM0070, AM0091, ACM0015, ACM0018). Other variants include “the average of the top five performers” (ACM0005, ACM0015), and a market penetration rate of 5% (AM0014, ACM0005).

The de-facto standard of “the average of the top 20% performer” is obviously derived from the baseline approach 48.c of the Marrakech Accords, a political decision made at COP7 in 2001. ACM0013 deviates from this trend and uses the average of top 15% performers. But this is only an exceptional political compromise made by the CDM EB after a heated debate on whether clean coal power projects should be eligible for the CDM at all. NM0302 attempted to use the top 20th and top 45th percentile levels for additionality demonstration and baseline setting respectively. The rejection of this methodology was largely due to the perception of the regulators that the proposed stringency levels were neither conservative enough nor justified with empirical data (UNFCCC, 2011c).

Stringency is one of the defining aspects of performance standard design. However, detailed technical judgments on the “right” level of stringency for the performance standard have not yet been developed, i.e., the stringency level that strikes an overall balance between over- and under-crediting of mitigation

efforts in a sector (Hayashi et al., 2010). This also explains why all the approved methodologies with a performance standard for additionality apply the same stringency level to baseline setting and additionality demonstration (AM0070, AM0091). Namely, as objective assessment of stringency levels is very difficult, these methodologies resorted to the politically decided top 20% average level for both baseline and additionality.

4.4 Data requirements

In most cases, performance standards are established based on empirical data from the most recent year (AM0014, AM0030, AM0037, AM0059, AM0063, AM0091, ACM0005, ACM0013, ACM0015). A longer data period is used only in a few methodologies: the most recent three years (AM0070, ACM0002, NM0302), and the most recent five years (AM0067). No projection-based data are used in the reviewed methodologies.

It is common to allow for the use of alternative data sources to empirical data. They include manufacturer's specifications (AM0063, AM0070), national regulations (AM0070, AM0091), and conservative default factors (AM0037, AM0063, AM0067, AM0070, ACM0002, ACM0005).

The data management system of NM0302 merits a detailed discussion. This methodology utilizes a global cement sector database of the Cement Sustainability Initiative. The database is managed by an independent third party, PricewaterhouseCoopers (PwC), in order to ensure confidentiality of business-sensitive information. In addition, a Project Management Committee serves as the single contact point for all communications between the database participants and PwC. Furthermore, the committee approves or rejects data queries submitted by stakeholders (CSI, 2012). Independent data management can thus

increase transparency in the performance comparison process, and safeguard the confidentiality of the collected data.

4.5 Updating procedures

The majority of the reviewed methodologies establishes a performance standard *ex ante*, and requires updating of the performance standard only at the renewal of a crediting period, i.e., every seven years (AM0037, AM0063, AM0067, AM0070, ACM0002, ACM0013, ACM0015, ACM0018).

Some methodologies require annual updating of performance standards by collecting actual data either as a mandatory option (AM0030, AM0059, AM0091, NM0302). For some, it is because the underlying database managed by a third party is anyway updated annually (e.g., International Aluminium Institute for AM0030 and AM0059; Cement Sustainability Initiative for NM0302). For others, annual updating is given as a voluntary option that is alternative to the *ex-ante* determination of a performance standard, and/or a default value for the annual updating is available (AM0070, ACM0002, ACM0005).

Most of the methodologies update all data sources for performance standard determination at the same frequency. AM0091 is the only exception that applies differentiated updating frequencies to the data sources. Namely, parameters that fluctuate greatly over time (e.g., building energy consumption) need more frequent updating than ones that are more stable (e.g., gross floor area).

5. Application of large-scale performance-standard methodologies to CDM projects

As of March 2012, there is a clear divide in the application of large-scale CDM methodologies using a performance standard. ACM0002 has been applied to a large number of projects, with 35% of all

registered projects applying this methodology. The rest of the approved methodologies accounts for just 1% of all registered projects. Among them, only ACM0005, ACM0013 and ACM0018 are used by more than 25 projects. But ACM0013 has been put on hold in late 2011 following allegations that projects covered by it are not additional, criticism of insufficiently frequent update of the performance standard and an overestimate of the emission reductions achieved (Lazarus and Chandler, 2011). This exemplifies the challenges performance standards are exposed to.

[Insert Table 3 here]

Data availability has been the main barrier to the use of performance standards. In consequence, the actual application of performance-standard methodologies has focused on sectors where a large body of data is already available (e.g., power). For example, the application of ACM0002 has greatly been facilitated by databases on power grid emission factors published by DNAs of CDM host countries. Almost 20 CDM host countries publish such databases on their DNA websites (Michaelowa, 2011).

Data availability is a particularly big concern for performance standards designed for dispersed end-use sectors, most notably for the building sector. Though no CDM project has been submitted applying AM0091, its design illustrates how data availability can become a key bottleneck for the building sector. The methodology defines a performance standard in specific emissions of a building unit (tCO_2/m^2). Building energy consumption data are the key to this performance standard. Such data are easier to obtain if the energy is supplied by local utilities and appropriate metering systems are implemented. This is likely the case with grid electricity supply and distribution of chilled/ hot water through a district system. However, if the energy is purchased or generated individually at individual building levels (e.g., LPG cylinder, central air conditioning), it would be challenging to collect energy consumption data from each building unit. In most developing countries, distributed building energy systems are far more common

than district ones. Thus, AM0091, as it currently stands, may well be workable only in advanced developing countries that have modern energy systems.

Another key reason for the limited use of these methodologies is the difficulty of proving additionality. Given that PFC reduction from aluminum production generally reduces production costs (AM0030, AM0059), and that the same applies for capture and use of gas in refineries (AM0037) as well as blending of cement with fly ash or slag (ACM0005, ACM0015), these project types have faced difficulties with additionality demonstration. Likewise, gas-based cogeneration plants are often commercially highly attractive (AM0014). Rejection rates for these four project types have therefore been high. The two methodologies that capture additionality through a performance standard, AM0070 and AM0091, have stringent monitoring requirements, and AM0091 is still very “young”. It thus remains to be seen whether they will be widely used, or whether further simplification is required.

6. Conclusions

The use of performance standards has emerged as an option to streamline baseline and additionality procedures of the CDM. While performance standards have been more popular for baseline determination, where over 10 methodologies have applied them, two performance standards have recently been approved for additionality demonstration for refrigerators and new buildings. There is so far a great variety in type of performance standards developed under the CDM. Although the emission or energy standard is the most common choice, there are also standards based on technology or practice specifications or market penetration rates. Despite this variety, the CDM regulators in 2011 principally approved only a smaller set of performance standards: emission standard, and technology or practice standard. In order for these performance standards to be applied, host country CDM authorities need to submit a standardized methodology proposal including the data used for its calculation.

Methodological approaches to performance standards have been elaborated greatly over time, especially on the aggregation aspect. However, the determination of performance standard stringency requires more sophisticated approaches than the de-facto standard of the Marrakech Accords (i.e., average of the top 20% performers) or the 90% (80%) stringency level agreed top-down in 2011 by the CDM EB. In order to elaborate updating procedures, detailed analysis is warranted on how emission or energy performance of a certain sector evolves over time. Also, differentiated updating frequencies of data parameters could help reduce monitoring costs.

Apart from the power sector, where necessary data are readily available, there has only been a limited number of projects actually applying the performance-standard methodologies. Data constraints in CDM host countries are the prime reason for the slow uptake on the ground. Thus, a careful attention should be given to balancing data requirements and practicability of performance standards. To this end, it would be helpful to allow for the use of nameplate parameters or default factors in case actual operating data are not available. Furthermore, local and/or international institutions with relevant expertise as well as industry associations could coordinate data collection efforts.

The above analysis shows that the devil is in the details of performance standard design, while recognizing the importance of ensuring user-friendliness of this approach. That is, performance standards should be designed as complex as necessarily, but be kept as simple as possible. The approaches used so far under the CDM tend to be situated at the simpler end. Thus, it is essential for CDM regulators to start opening the black box when further standardizing CDM methodologies.

References

Broekhoff, D., 2007, *Expanding global emissions trading: Prospects for standardized carbon offset crediting*, International Emissions Trading Association, Geneva.

CSI, 2011, *CO₂ and energy accounting and reporting standard for the cement industry: The cement CO₂ and energy protocol*, Cement Sustainability Initiative, Geneva.

CSI, 2012, *Global cement database on CO₂ and energy information*, Cement Sustainability Initiative, Geneva [available at <http://wbcsdcement.org/index.php/key-issues/climate-protection/global-cement-database>].

Gillenwater, M., 2011, *What is Additionality? Part 1: A long standing problem*, GHG Management Institute Discussion Paper 1, Silver Springs.

Hayashi, D., Michaelowa, A., 2007, *Lessons from submission and approval process of large-scale energy efficiency CDM methodologies*, HWWI Research Paper No.4-11, HWWI, Hamburg.

Hayashi, D., Müller, N., Feige, S., Michaelowa, A., 2010, *Towards a more standardized approach to baselines and additionality under the CDM*, Perspectives, Zurich.

IEA, 2007a, *World energy outlook 2007: China and India insights*, International Energy Agency, Paris.

IEA, 2007b, *Tracking industrial energy efficiency and CO₂ emissions*, International Energy Agency, Paris.

IETA, 2009, *Multi-project, standardized baselines: Explaining a key issue in the reform of the Clean Development Mechanism*, International Emissions Trading Association, Geneva [available at www.ieta.org/assets/Reports/ieta-%20standardized%20baselines%20explained.pdf].

Kartha, S., Lazarus, M., LeFranc, M., 2005, 'Market penetration metrics: Tools for additionality assessment?', *Climate Policy* 5, 147-165.

Lazarus, M., Kartha, S., Ruth, M., Bernow, S., 1999, *Evaluation of benchmarking as an approach for establishing Clean Development Mechanism baselines*, Tellus Institute/Stockholm Environment Institute, Boston, MA.

Lazarus, M., Kartha, S., Bernow, S., 2000, *Key issues in benchmark baselines for the CDM: Aggregation, stringency, cohorts, and updating*, Tellus Institute/Stockholm Environment Institute, Boston, MA.

Lazarus, M., Chandler, C., 2011, *Coal power in the CDM: Issues and options*, Stockholm Environmental Institute, Stockholm.

Michaelowa, A., 2005, 'Determination of baselines and additionality for the CDM: A crucial element of credibility of the climate regime'. in: F. Yamin (ed), *Climate change and carbon markets: A handbook of emission reduction mechanisms*, Earthscan, Oxford, 289-304.

Michaelowa, A., 2009, 'Interpreting the additionality of CDM projects: Changes in additionality definitions and regulatory practices over time', in: D. Freestone, C. Streck (eds), *Legal Aspects of Carbon Trading*, Oxford University Press, Oxford, 248-271.

Michaelowa, A., 2011, *Rule consistency of grid emission factors published by CDM host country authorities*, CDM Watch, Brussels [available at www.cdm-watch.org/wordpress/wp-content/uploads/2011/02/rule_consistency_of_grid_emission_factors_published_by_CDM_host_country_authorities_14_Feb_2011-.pdf].

Murtishaw, S., Sathaye, J., LeFranc, M., 2006, 'Spatial boundaries and temporal periods for setting greenhouse gas performance standards', *Energy Policy* 34, 1378-1388.

Neelis, M., Worrell, E., Müller, N., Angelini, T., Cremer, C., Schleich, J., Eichhammer, W., 2009, *Developing benchmarking criteria for CO₂ emissions*, Ecofys/ Fraunhofer Institute for Systems and Innovation Research, Utrecht/ Karlsruhe.

URC, 2012, *CDM pipeline (1 March 2012 update)*, UNEP Risø Centre, Roskilde [available at www.cdmpipeline.org].

UNFCCC, 1998, *Kyoto Protocol*, United Nations Framework Convention on Climate Change, Bonn.

UNFCCC, 2001, *Decision 17/CP.7: Modalities and procedures for a Clean Development Mechanism, as defined in Article 12 of the Kyoto Protocol*, FCCC/CP/2001/13/Add.2, United Nations Framework Convention on Climate Change, Bonn.

UNFCCC, 2006, *Decision 3/CMP.1: Modalities and procedures for a clean development mechanism as defined in Article 12 of the Kyoto Protocol*, FCCC/KP/CMP/2005/8/Add.1, United Nations Framework Convention on Climate Change, Bonn [available at <http://cdm.unfccc.int/Reference/COPMOP/08a01.pdf>].

UNFCCC, 2009, *Executive Board of the Clean Development Mechanism forty-ninth meeting report*, United Nations Framework Convention on Climate Change, Bonn [available at <http://cdm.unfccc.int/EB/049/eb49rep.pdf>].

UNFCCC, 2010, *Standardized baselines under the clean development mechanism*, FCCC/TP/2010/4, United Nations Framework Convention on Climate Change, Bonn [available at <http://unfccc.int/resource/docs/2010/tp/04.pdf>].

UNFCCC, 2011a, *Decision 3/CMP.6: Further guidance relating to the Clean Development Mechanism*, FCCC/KP/CMP/2010/12/Add.2, United Nations Framework Convention on Climate Change, Bonn [available at <http://unfccc.int/resource/docs/2010/cmp6/eng/12a02.pdf#page=2>].

UNFCCC, 2011b, *CDM EB65 report annex 23: Guidelines for the establishment of sector specific standardized baselines*, United Nations Framework Convention on Climate Change, Bonn [available at http://cdm.unfccc.int/Reference/Guidclarif/meth/meth_guid42.pdf].

UNFCCC, 2011c, *Meth Panel final recommendation on NM0302 “Emission reductions in the cement production facilities of Holcim Ecuador S.A.*, United Nations Framework Convention on Climate Change, Bonn [available at <http://cdm.unfccc.int/filestorage/C/T/F/CTFSBGVQHMPE0981D2K7I4A5LOJN3U/Final%20recommendation%20for%20NM0302.pdf?t=bjN8bTZhNHRhfDCdPu5E5vSAEcYDLngmZSfy>].

UNFCCC, 2012, *CDM EB63 report annex 28: Procedure for submission and consideration of standardized baselines (version 01.0)*, United Nations Framework Convention on Climate Change, Bonn [available at http://cdm.unfccc.int/Reference/Procedures/meth_proc07.pdf].

Table 1: Principles of CDM methodologies in the context of standardized approaches

Principle	Description
Objectivity	Avoid subjective judgment in the design and implementation of a methodology.
Applicability	Ensure applicability of a methodology in a wide set of geographical locations and project types.
Usability	Design a methodology as simply as possible so that it can be used in an accessible way.
Consistency	Make sure methodological procedures are internally consistent, and are in line with procedures in other relevant approved methodologies.
Transparency	Develop a methodology without resorting to black-box procedures so that it can be implemented in a transparent manner.
Predictability	Maximize predictability in the estimation and monitoring of emission reductions.
Environmental integrity	Keep an overall (not project-specific) balance between over- and under-crediting of emission reductions.

Source: Authors based on UNFCCC (2006 and 2009).

Table 2: Reviewed CDM methodologies with performance standard application

Name	Version	Sector
AM0014 “Natural gas-based package cogeneration”	4	Power
AM0030 “PFC emission reductions from anode effect mitigation at primary aluminum smelting facilities”	3	Aluminum
AM0037 “Flare (or vent) reduction and utilization of gas from oil wells as a feedstock”	2.1	Oil & gas
AM0059 “Reduction in GHGs emission from primary aluminum smelters”	1.1	Aluminum
AM0063 “Recovery of CO ₂ from tail gas in industrial facilities to substitute the use of fossil fuels for production of CO ₂ ”	1.2	Any industry
AM0067 “Installation of energy efficient transformers in a power distribution grid”	2	Power
AM0070 “Manufacturing of energy efficient domestic refrigerators”	3.1	Building
AM0091 “Energy efficiency technologies and fuel switching in new buildings”	1	Building
ACM0002 “Consolidated baseline methodology for grid-connected electricity generation from renewable sources” ¹	12.3	Power
ACM0005 “Increasing the blend in cement production”	6	Cement
ACM0013 “New grid connected fossil fuel fired power plants using a less GHG intensive technology”	4	Power
ACM0015 “Project activities using alternative raw materials that do not contain carbonates for clinker production in cement kilns”	3	Cement
ACM0018 “Electricity generation from biomass residues in power-only plants”	1.3	Power
NM0302 “Cement and clinker production facilities based on benchmarking”	2	Cement

¹ In October 2007, the performance standard approach in ACM0002 was incorporated into “Tool to calculate the emission factor for an electricity system”. However, the table lists ACM0002 as it is the original source of the performance standard approach.

Source: Authors.

Table 3: Utilization of CDM methodologies with performance standard application

Meth approval date	Name	Sector	Number of projects			Pre-2013 CER volume (million) ¹
			Registered	At validation	Rejected	
Sep. 2004	AM0014	Power	6	0	6	0.7
Sep. 2004	ACM0002	Power	1,361	1,399	76	800.1
Sep. 2005	ACM0005	Cement	14	33	23	23.5
May 2006	AM0030	Aluminum	3	0	2	0.7
Sep. 2006	AM0037	Oil & gas	2	0	4	0.5
Oct. 2007	AM0059	Aluminum	1	2	0	2.6
Sep. 2007	ACM0013	Power	6	36	7	21.5
Nov. 2007	AM0063	Any industry	0	2	0	0.1
Nov. 2007	ACM0015	Cement	0	5	6	1.0
Mar. 2008	AM0067	Power	0	0	0	0.0
Sep. 2008	AM0070	Building	0	2	0	0.5
Feb. 2010	ACM0018	Power	5	39	1	5.1
Jun. 2011	AM0091	Building	0	0	0	0.0

¹ Counting CERs from projects that are registered or at validation.

Source: Authors based on URC (2012).

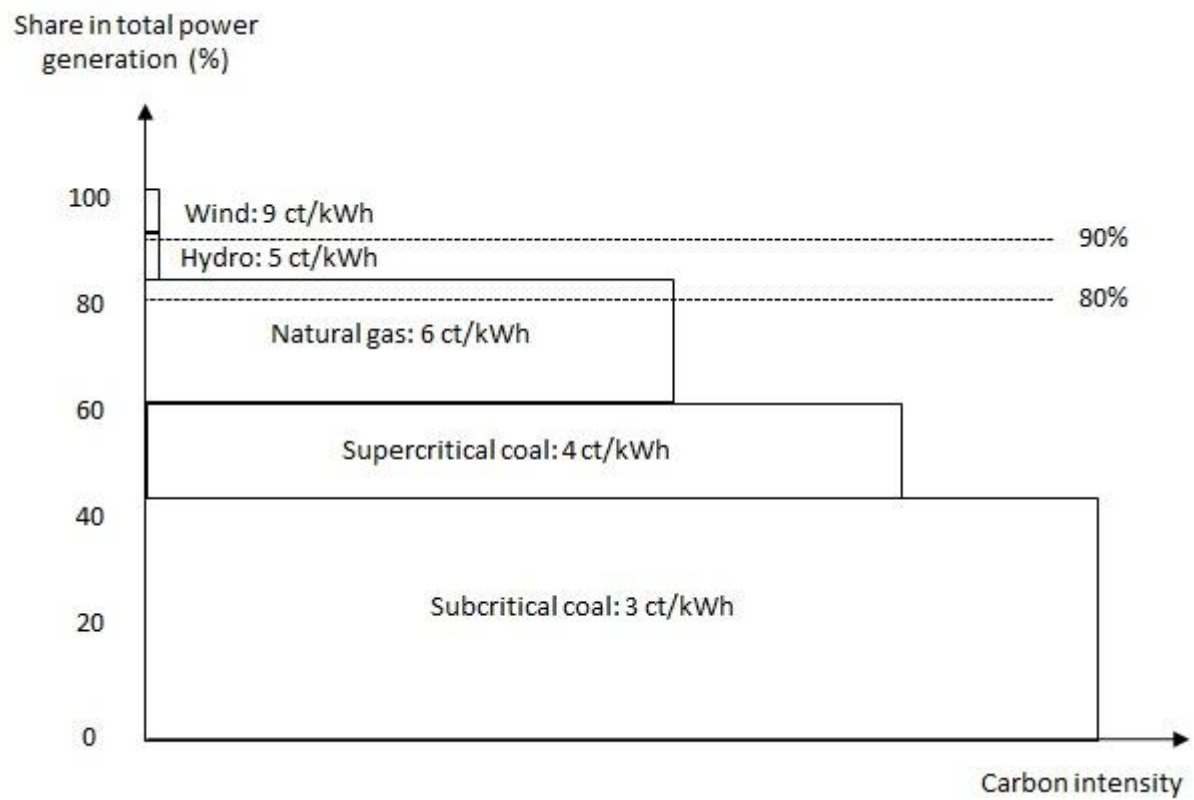


Figure 1: Emission performance standard for electricity generation as per CDM EB guideline

Source: Authors.

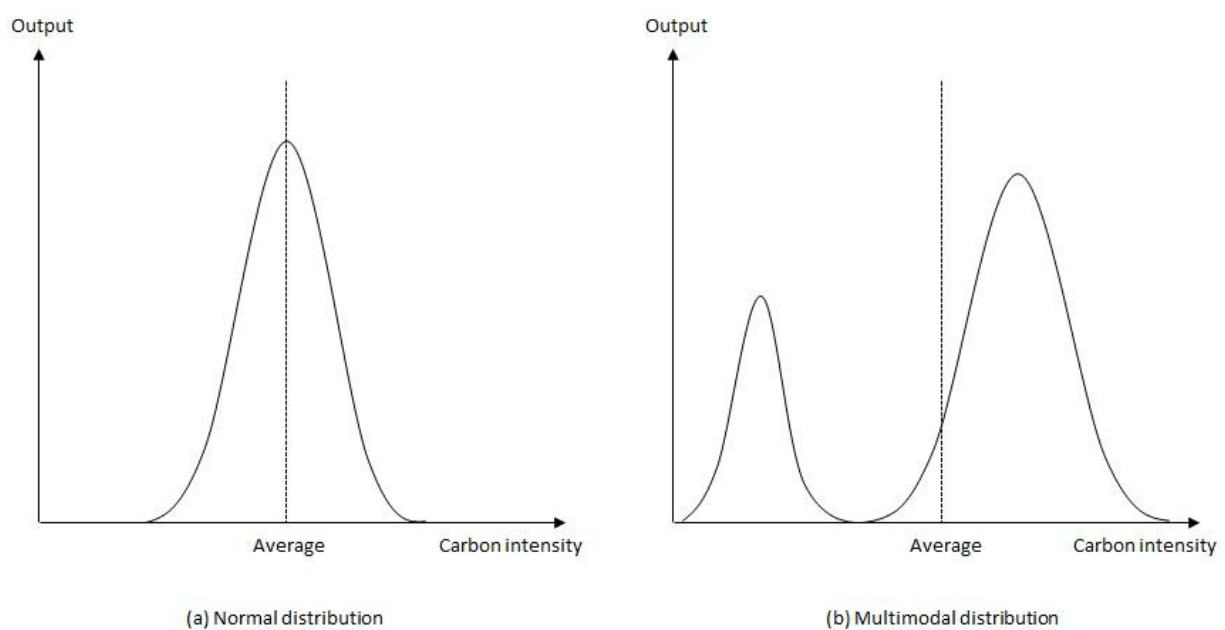


Figure 2: Stringency level for normal and multimodal performance distribution

Source: Authors.

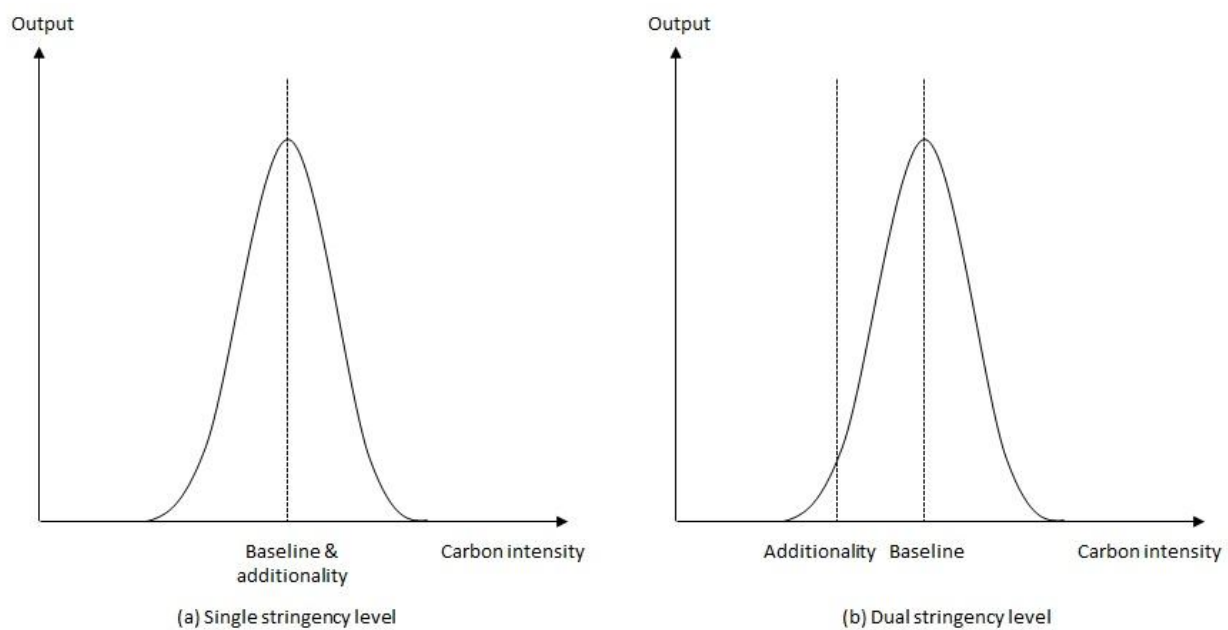


Figure 3: Single or dual stringency level for baseline and additionality

Source: Authors.